

University of Dundee

Balancing measures or a balanced accounting of improvement impact

Toma, Madalina; Dreischulte, Tobias; Gray, Nicola M; Campbell, Diane; Guthrie, Bruce

Published in:
BMJ Quality & Safety

DOI:
[10.1136/bmjqs-2017-006554](https://doi.org/10.1136/bmjqs-2017-006554)

Publication date:
2018

Document Version
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Toma, M., Dreischulte, T., Gray, N. M., Campbell, D., & Guthrie, B. (2018). Balancing measures or a balanced accounting of improvement impact: a qualitative analysis of individual and focus group interviews with improvement experts in Scotland. *BMJ Quality & Safety*, 27(7), 547-556. <https://doi.org/10.1136/bmjqs-2017-006554>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Balancing measures or a balanced accounting of improvement impact. A qualitative analysis of individual and focus group interviews with improvement experts in Scotland

ABSTRACT

Background

As quality improvement (QI) programmes have become progressively larger-scale, the risks of implementation having unintended consequences is increasingly recognised. More routine use of balancing measures to monitor unintended consequences has been proposed to evaluate overall effectiveness, but in practice published improvement interventions hardly ever report identification or measurement of consequences other than intended goals of improvement.

Methods

We conducted 15 semi-structured interviews and two focus groups with 24 improvement experts to explore the current understanding of balancing measures in QI and inform a more balanced accounting of the overall impact of improvement interventions. Data were analysed iteratively using the framework approach.

Results

Participants described the consequences of improvement in terms of desirability/undesirability and the extent to which they were expected/unexpected when planning improvement. Four types of consequences were defined: expected desirable consequences (*goals*); expected undesirable consequences (*trade-offs*); unexpected undesirable consequences (*unpleasant surprises*) and unexpected desirable consequences (*pleasant surprises*). Unexpected consequences were considered important but rarely measured in existing programmes, and an improvement pause to take stock after implementation would allow these to be more actively identified and managed. A balanced accounting of all consequences of improvement interventions can facilitate staff engagement and reduce resistance to change, but has to be offset against the cost of additional data collection.

Conclusion

Improvement measurement is usually focused on measuring intended *goals*, with minimal use of balancing measures which when used, typically monitor *trade-offs* expected before implementation. This paper proposes that improvers and leaders should seek a balanced accounting of all consequences of improvement across the life of an improvement programme, including deliberately pausing after implementation to identify and quantitatively or qualitatively evaluate any *pleasant* or *unpleasant surprises*.

Keywords

Quality improvement; Measurement of quality; Balancing measures; Unintended consequences; Expected, unexpected, desirable and undesirable consequences

Word count for the abstract: 275

Word count for the text of the manuscript: 4000

BACKGROUND

Unintended consequences with negative or positive effects on care processes and outcomes can occur with any change in complex systems like healthcare organisations,¹⁻³ and so are an important potential problem in quality improvement (QI).⁴⁻⁶ More routine use of balancing measures to account for and manage unintended consequences of improvement interventions is recommended by a number of organisations.⁷⁻¹⁰ The Institute of Healthcare Improvement (IHI) for example describes measurement in improvement programmes in terms of process and outcome measures focused on delivering pre-defined intended benefits, and balancing measures in terms of negative unintended consequences in other parts of the healthcare system (Box 1).^{7,8} Reflecting this perspective, hospital readmission rates are often used as a balancing measure for interventions aiming to reduce the length of hospital stay, since it is plausible that shortening length of stay could mean discharging patients who are then unable to manage at home.¹¹⁻¹³

Despite calls for a more systematic accounting of all side effects of improvement interventions,^{14,15} a number of systematic reviews have shown that balancing measures appear rarely used or reported in practice. A review of the application of Plan Do Study Act (PDSA) methods found that only six (6.4%) of 94 included studies reported any “*disconfirming observations*” about the intervention,¹⁶ and only one of 100 included studies in a systematic review of perioperative care improvement interventions reported an “*unfavourable or unintended sign, symptom or event*”.¹⁷ These findings are consistent with other reviews, including one of the application of improvement methodologies in surgery which found that none of 34 included studies reported on unintended consequences,¹⁸ and another where only one of 121 studies of interventions to reduce patient falls and catheter-associated infections measured any unintended consequences.¹⁹ Several other studies in the latter review provided anecdotal evidence of “*unexpected occurrences*”,¹⁹ but robust evaluation of such claims is rare in improvement programmes more generally.²⁰ There is additionally little evidence that improvers routinely consider the potential for unexpected consequences post-implementation,²¹ and the amount of missing data about outcomes other than goals is often significant.^{22,23} The aim of this paper is to explore current understanding of balancing measures in healthcare improvement, including the range of consequences that could, or should be considered to inform a more balanced accounting of the overall impact of improvement interventions.

METHODS

Design and participants

The research was carried out in two phases, with semi-structured interviews used in the initial phase to formulate a draft conceptual framework for considering all consequences of improvement which was then explored using focus group interviews to refine and elaborate the framework, and to consider its wider applicability.

We used purposive sampling to include a broad spectrum of stakeholders with expertise in metrics and measure design in healthcare QI or relevant clinical and/or academic experience in improvement implementation. Participants in both phases of the study included improvement advisors, clinical academics, providers of health and social care services, policy-makers and patient representatives identified from relevant publication records and major conferences on QI, members of QI groups, on-line searches of open-access information and research teams' networks and contacts. Participants were largely based in Scotland, where comprehensive healthcare, which is free at the point of care, is provided to all residents by the taxpayer-funded National Health Service (NHS). Digital maturity of the system varies, with all primary care practices exclusively using electronic medical records (EMR) with widespread electronic data sharing (including for example, primary care sharing of data for hospital use in an emergency care summary, electronic transmission of letters and discharge summaries, and automated laboratory results transmission), but hospitals being at various stages of EMR implementation. NHS Scotland has invested significantly in staff training in improvement and introduced a number of centrally led national safety and quality improvement programmes²⁴, largely (but not exclusively) based on the IHI Model for Improvement. Additional participants with particular expertise or known interest in measurement were purposively recruited from England and the United States. All participants were actively involved in service improvement across various settings including social care, mental health, public health, medicine for the elderly, maternity, neonatal and paediatric care.

Data collection and analysis

Phase 1 -Semi-structured interviews to formulate the framework

Twelve face to face semi-structured interviews and three telephone interviews each lasting for approximately one hour explored participants' understanding of balancing measures as part of a broader discussion about QI methods in health and social care. Individual interviews followed a topic guide based on the published literature and two pilot interviews. Data were analysed according to the principles of the Framework approach²⁵ by developing codes and categories from the transcripts

and grouping them into a preliminary coding matrix. The Diffusion of Innovation literature²⁶⁻³⁰ was used to reinterpret the initial matrix and generate a more structured framework reflecting participants' conceptualisation of balancing measures. The researcher who conducted the interviews (MT) coded all transcripts with a selection of transcripts and the emerging framework reviewed by a second experienced researcher (BG) to refine the coding.

Phase 2 – Focus groups interviews to refine the framework

Two focus groups were conducted to explore the current understanding of balancing measures in QI and to elaborate the framework generated in Phase 1. The draft framework was shared in a briefing paper prior to the focus group meeting and was used to inform initial discussions within the groups. Focus groups were facilitated by two experienced moderators (MT and BG), lasted about 75 minutes each, and took place on a single day. Interviews and focus groups were conducted in a non-directive manner, with participants encouraged to talk openly and with relative freedom to steer the discussion. The main researcher kept a journal with field notes reflecting on the research process, including prior assumptions that might have influenced the findings. Data were analysed using an iterative and step-wise process. The framework developed in phase one was used as a coding matrix in the analysis. Codes from focus groups transcripts were grouped into sub-themes, which were then allocated to one of the domains of the initial framework. One researcher (MT) coded all data and the wider team met regularly to reach consensus on the final framework structure, discuss additional categories, and resolve any disagreement.

All interviews and focus group data were audio-recorded, transcribed verbatim and analysed using NVivo11.

FINDINGS

Participants' characteristics

Semi-structured interviews with 15 participants and two focus groups with 24 participants (two of whom were also interviewed in phase 1) were completed. Participants had a wide range of roles in improvement and implementation science. 32 participants came from Scotland, four from England and one from the United States. (Table 1).

Phase 1-Semi-structured interviews

Identifying key themes and concepts

When asked about their overall understanding of balancing measures, participants initially emphasised negative consequences of improvement in other parts of the healthcare system, paralleling the IHI definition.

“My understanding is that a balancing measure is essentially something that you put in place because you recognise that often you can go in with the best of intent to improve an issue, you can deliver the improvement but you just end up creating more problems somewhere else” (Improvement advisor)

Specific examples were again typically framed negatively, often as “adverse” or “knock-on” effects. Some of these were described as predictable from the outset, and measured routinely in the local improvement context.

“The mental health safety programme has balancing measures around recovery, about being very clear that one way of improving safety could lead to less positive risk-taking, which would be a very negative unintended consequence. We always use the Scottish recovery indicator, making sure that we promote recovery-oriented practices and we're not clamping down on folk.” (Mental health care provider)

Other negative consequences were described as only emerging as a potential problem after initial implementation, requiring improvers to be sensitive to the possibility of harm, and to be ready to ask themselves “right, what are we going to put in place to measure these adverse effects and see whether the improvement is actually causing any harm?” (Academic and public health specialist), in order to inform further investigation or action.

“Work to increase rates of early discharge and reduce length of stay led to patients being discharged into inappropriate conditions which in turn caused an increase in costs and readmission rates (...) That should be a wee bit of a red flag for you to think ‘why is everybody coming back? Are they coming back in because of surgical site infections or because you didn’t get their medicines reconciliation right on discharge?’ (...)” (Improvement advisor)

Less commonly, participants described unanticipated positive or beneficial consequences. Although they were often uncertain whether these could be considered ‘balancing measures’ since they did not balance the benefits of improvement, they were highly valued by those who had experience of them.

165 *“A QI initiative aimed at improving writing and reading skills in secondary schools led to a*
166 *reduction in absence rates as a result of better students’ engagement with different activities*
167 *across the school (...) It was actually quite surprising and certainly a delightful outcome that*
168 *we can now flip into a new piece of work to support children to become more engaged across*
169 *their whole learning journey.” (Provider of social care services)*

170 However, in practice, the use of balancing measures was perceived to be rare in large-scale
171 healthcare improvement programmes.

172 *“Most safety programmes haven't paid much attention to balancing measures. From forty-*
173 *nine pages of measures [in a safety improvement programme], there's probably only two or*
174 *three balancing measures like readmission rates, average length of stay or reintubation rates*
175 *when reducing the time patients spend on a ventilator after surgery (...)” (Policy maker*
176 *health and social care)*

177 **Formulating the framework**

178 In summary, when first asked about balancing measures, participants typically started from the
179 position that measures should be implemented to assess undesirable unintended consequences of
180 improvement work. However, their subsequent description of balancing measures also included
181 unanticipated desirable consequences, and considerable discussion of the extent to which all
182 consequences were predictable from the outset. Drawing on the Diffusion of Innovation literature,²⁶⁻
183 ³⁰ we developed an initial framework that describes the range of consequences that improvement
184 could have, in terms of their desirability and the extent to which they were anticipated when
185 planning improvement.

186 Four types of consequence were defined at this stage and described as *goals, trade-offs, classic*
187 *negative unintended consequences and serendipities*. (Figure 1, sent to phase 2 participants before
188 the focus groups)

189 **Phase 2-Focus group interviews**

190 **Mapping key themes and concepts**

191 Similar to the individual interviews, focus group participants initially described balancing measures in
192 terms of trade-offs, i.e. negative unintended consequences of QI that were expected from the
193 outset.

194 *“A lot of potential consequences are known at the start. ‘Oh, we need to actually count that,*
195 *it will be an interesting balancing measure’. In a recent project focused on improving growth*
196 *by early enteral feeding and maximise use of parenteral nutrition, the rates of necrotising*

197 *enterocolitis and community-acquired bloodstream infections had reasonable potential for a*
198 *balancing measure.” (Provider of neonatology services)*

199 However, as in the individual interviews, participants discussed several examples when undesirable
200 consequences only became apparent after implementation, with examples from the same area of
201 care targeted by improvement, as well as other parts of the wider system.

202 *“Inducing pregnant women at 40 weeks aimed to decrease the risk of stillbirth and newborn*
203 *death but led to the use of extra interventions such as continuous fetal monitoring (...) which*
204 *in turn increased costs and decreased overall patient satisfaction. Also woman who had a*
205 *serious medical need for an induction could not get on the schedule because all of the*
206 *hospital beds were occupied by women being electively induced.” (Provider of maternal and*
207 *infant healthcare)*

208 Participants also mentioned desirable unintended consequences referring to *“serendipitous side*
209 *effects or bonuses which are not planned as original programme outcomes”* (academic and primary
210 care provider), which they said were important to consider in order to obtain a balanced view of the
211 overall impact of improvement interventions.

212 *“The Book Bug sessions were established to strengthen attachment between parents and*
213 *children by encouraging them to share and enjoy books together. One of the measures,*
214 *which wasn't a balancing measure in the first instance but turned into one, was an increased*
215 *interest from parents to improve their own literacy, bearing in mind that they had a young*
216 *child that would need supported through school.” (Public health specialist)*

217 However, even when unintended consequences were clearly identified, concerns were raised about
218 the difficulty of creating or implementing a fully balanced set of measures, since data was not
219 usually available from the outset unless routinely accessible from an existing source.

220 *“I think we struggle with balancing measures. We always know we should think about them*
221 *beforehand, but don't know how to deal with what comes up during the project (...) I think in*
222 *safety we probably talk more about negative expected consequences, and the unexpected*
223 *ones are the tip of the iceberg stuff (...) I don't think we become aware of them very often*
224 *and we tend to then think ‘oh it would have been nice to have data on that at the beginning’.*
225 *(...) they almost feel like a missed opportunity.” (Academic capacity building)*

Barriers and facilitators to using balancing measures

In terms of measure design, the majority of interviewees found the distinction between ‘process’, ‘outcome’ and ‘balancing’ measures in some of the improvement literature confusing, since balancing measures could relate to processes and outcomes depending on the context.

“We tend to be quite prescriptive about the family of measures and putting things into baskets of process and outcome and balancing measures is not always helpful. I don’t think we pay enough attention to balancing measures and I’m not sure whether they’re the right ones either (...) Readmission rates and average length of stay are balancing measures, but they could also be outcomes or processes that we might measure.” (Academic and palliative care provider)

Participants broadly perceived balancing measures to be important and relatively underused but reflected on the increasing burden of data collection in already resource-constrained systems.

“The time that we spend collecting or looking for data is time we don’t spend delivering patient care, so there’s a cost to this. Having balancing measures could be disproportionately expensive (...) just one of those things when measures are added on and on and nothing’s changing. You’re just collecting for the sake of collecting. You need to consider these measures very carefully or it’s a waste of peoples’ time.” (Provider of geriatric healthcare)

However, there was a general agreement that engaging those involved in delivering care in the choice and design of measures from the outset would likely lead to better understanding of the rationale for measuring and could help minimise the burden of data collection.

“If the work is owned by the frontline staff, if it’s their piece of improvement and if they’ve developed their own balancing measures then they’re not going to think that measurement is too onerous in the same way as other would if they don’t understand why they’re measuring.” (Policy maker education and early years)

More importantly, the overall process of considering unintended consequences and implementing balancing measures was perceived to have value in its own right in terms of improving staff engagement with improvement and overcoming resistance to change.

“You find a lot of latent resistance because people are genuinely worried about an unintended consequence and they don’t engage in the work. You can introduce your checklist and it is fantastic, but it really annoys the staff because “this is just going to take up a huge amount of time’ (...) Using a balancing measure can convince your communities that

improvement is needed and could be a goodwill builder if people know that you're monitoring and taking their concerns seriously.” (Academic community engagement)

Refining the framework

Figure 2 shows a revised version of the framework that takes account of focus groups findings, including the language used (eg ‘expected’ rather than ‘anticipated’). Desirability was described as a clear dichotomy, but expectations were perceived as more of a spectrum. While an initial measurement plan can define consequences expected from the outset (*goals* and *trade-offs*), participants thought that improvement programmes might need to plan for a ‘pause’ after implementation to account for unexpected consequences, both desirable and undesirable. The language of ‘serendipities’ and ‘classic negative unintended consequences’ was disliked, and renamed. The four type of consequences in the revised framework (Figure 2) were therefore: **Improvement goals**: the expected and desirable consequences of the improvement programme, defined by the initial measurement plan; **Improvement trade-offs**: the expected but undesirable consequences of the improvement programme, and implicitly believed to be smaller in magnitude than the goals (and so an acceptable compromise); **Pleasant surprises**: unexpected and desirable consequences emerging after implementation; **Unpleasant surprises**: unexpected and undesirable consequences emerging after implementation.

All four consequences can be measured using either process or outcome measures and can arise in the same area of care targeted by improvement, or elsewhere in the health and social care system.

DISCUSSION

Summary of findings

Participants started by discussing balancing measures in terms of undesirable consequences which were expected before or early in implementation (*trade-offs*) and which could offset some of the intended benefits of improvements (*goals*). Although a range of examples were discussed, most participants agreed that such measures were relatively rarely used. Participants additionally emphasised that many consequences only became apparent after implementation, and these unexpected consequences could be either desirable or undesirable (*pleasant* or *unpleasant surprises*) and could accrue in the same part of the system as the improvement work, or other parts. There was frequent confusion as to what a balancing measure should measure, since the implication of many existing framings^{7 8} is that balancing measures are distinct in some way from process and outcome measures, rather than any type of consequence being measurable in terms of processes and outcomes. Involving front-line staff in identifying unintended consequences and balancing measure design was perceived to increase engagement with improvement and reduce resistance to

change. Balancing measures were seen as a necessary and integral part of evaluating the impact of an improvement programme, as well as a pragmatic way of engaging sceptics constructively by understanding their legitimate concerns around implementation. However, the value of designing and implementing balancing measures has to be offset against their cost in the context of overall measurement burden.

Strengths and limitations of the study

A strength of the study is that it drew on both empirical data from a purposively wide range of stakeholders and existing literature on unintended consequences. A limitation is that the sample was largely recruited from Scotland which may limit generalisability. However, NHS Scotland has a history of centrally led, and broadly successful efforts to introduce system-wide improvement interventions, most commonly based on the IHI Model for Improvement including training and implementation of national safety programmes in acute hospitals, mental health care and primary care.²⁴ Participants therefore had experience of a number of improvement programmes to draw on, although limited implementation of electronic medical records in hospitals means that perceptions of the burden of data collection will at least partly reflect that data used in national improvement programmes currently almost entirely consists of bespoke data collected by clinical staff. Findings were consistent across the diverse range of stakeholders (including those outside in Scotland), and we believe that the measurement issues faced by improvement programmes in Scotland are likely to be relevant in other countries and systems worldwide.

Comparison with existing literature

The existing improvement literature on measurement design emphasises the importance of developing a balanced set of measures during the planning of an improvement programme,^{7-10 31-33} often distinguishing between process and outcome measures for goals, and balancing measures for expected undesirable consequences (*trade-offs*) which are easily predictable from the outset (Box 1). However, participants in this study found this framing too narrow because they were concerned about unexpected undesirable consequences (*unpleasant surprises*) and valued unexpected desirable consequences (*pleasant surprises*), neither of which could be defined prior to intervention implementation.

Although there are some studies of *trade-offs*,³⁴⁻³⁶ and *pleasant*³⁷ and *unpleasant*^{38 39} *surprises* (Table 2), published improvement interventions rarely report data relating to unintended consequences.^{15-19 40} This may partly reflect publication bias, since authors are known to emphasise positive results and “tuck away”⁴¹ negative contextual features and failures.²³ However, it also likely reflects more general lack of consideration or measurement of unintended consequences, consistent with an

observed preoccupation with measuring pre-specified local processes and outcomes (*goals*).⁴²⁻⁴⁴ The implementation of PDSA cycles in healthcare for example has been criticised for often involving an over-simplified “*Do, Do, Do*” approach¹⁵ focused on little and often measurement and delivery of *goals* at the expense of thinking ahead and looking to the future (for *trade-offs*) and reflecting on potential hazards during implementation (for *surprises*).^{45 46}

Implications for improvement programme design

Balancing measures are an integral and core element of commonly used improvement models like the IHI Model for Improvement,^{7 8} but they are sometimes poorly specified and do not appear to be commonly implemented in practice.^{15-19 40} Based on the literature and the findings of this study, we believe that rather than focusing on balancing measures to implement at the start of improvement, improvers and leaders at all levels of management should consider how best to achieve a balanced accounting of the overall impact of improvement across the life of a programme. This requires consideration of all four types of consequence, any of which can be measured in terms of process and outcome. (Figure 2) Such a balanced accounting of impact can be achieved by articulating clear assumptions and formulating explicit predictions for both *goals* and *trade-offs* before implementation,^{14 40 47} and having a planned improvement pause after implementation to deliberately step back from *goal* delivery to take stock and reflect on potential *surprises*.^{46 48} In an ideal world, improvers would consult the available evidence base and seek external input from key stakeholders in order to identify potential *trade-offs*, speculate on and investigate potential *surprises*, and if necessary, to design relevant process and outcome measures to account for them. However, improvement takes place in resource-constrained environments, which will confine what is possible, including for example, the feasibility of measurement in other areas of a complex system. Focusing on a balanced accounting rather than balancing measures also emphasises that qualitative methods have much to offer both for the identification of *trade-offs* before implementation, and for understanding *surprises* after implementation where retrospective measurement may be difficult.⁴⁹

⁵⁰

Implications for reporting quality improvement projects

Few improvement reports mention unintended consequences, despite the Standards for Quality Improvement Reporting Excellence (SQUIRE) guidance¹⁴ including a requirement that reporting should include “*unintended consequences, such as unexpected benefits, problems, failures or costs associated with the intervention*” (standard 13e). Of note though is that the SQUIRE explanation and elaboration for this standard⁵¹ focuses more on exploring variation in implementation effectiveness and does not provide any examples of significant elaboration of unintended consequences. As the volume of publications in QI is growing, modification of SQUIRE to clarify that improvement reports

357 *should* report any measured or qualitatively assessed unintended consequences, or report that these
358 were not assessed, would be helpful to contextualise any evidence presented about the
359 achievement of improvement goals.

360 **CONCLUSION**

361 This study is largely based on analysis of data from interviews carried out in Scotland which has an
362 integrated single-payer healthcare system and relatively well-developed quality improvement
363 infrastructure.⁵² However, improvement interventions in complex systems will often result in
364 unintended consequences irrespective of context, so we believe that the conclusions apply more
365 widely, although the ability of improvers to evaluate or measure unintended consequences will vary,
366 being lower in more fragmented healthcare systems. Overall, the evidence is that improvement
367 programme measurement is usually focused on evaluating intended *goals*, with minimal use of
368 balancing measures which are typically monitoring *trade-offs* expected before implementation. We
369 conclude that a more balanced accounting of the effects of improvement should consider *goals* and
370 predictable *trade-offs* early in the design of an improvement programme, and also pause to take
371 stock of *pleasant and unpleasant surprises* after a period of implementation.

372 **List of abbreviations**

373

374 **QI:** Quality improvement

375 **IHI:** Institute of Healthcare Improvement

376 **NHS:** The National Health Service

377 **EMR:** Electronic Medical Record

378 **PDSA:** Plan Do Study Act

379 **SQUIRE:** Standards for Quality Improvement Reporting Excellence

380

381 **DECLARATIONS**

382

383 **Acknowledgements**

384 This work was undertaken by and on behalf of The Scottish Improvement Science Collaborating
385 Centre (SISCC). We are grateful to our improvement experts for their time and willingness to co-
386 create the “*balanced accounting*” framework. We also acknowledge Julie Anderson and Anita Lee,
387 who provided invaluable feedback on the revised versions of this paper.

388 **Ethics approval and consent to participate**

389 The University of Dundee Research Ethics Committee waived ethical approval for this study (UREC
390 15069). Informed consent to participate in the study was obtained from all participants.

391 **Consent for publication**

392 Not applicable

393 **Availability of data and material**

394 The data used and/or analysed during the current study are available from the corresponding author
395 on reasonable request.

396 **Competing interests**

397 The authors declare that they have no completing interests.

398

Funding

This paper presents core work from the Scottish Improvement Science Collaborating Centre, funded by the Scottish Funding Council, Health Foundation, Chief Scientist Office, and NHS Education Scotland with in-kind contributions from participating partner universities and health boards. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript

Authors' contributions

MT and BG were responsible for planning the study and led the data collection and analysis. TD, NG and DC contributed to data analysis. MT drafted and led the writing of the manuscript. BG, TD, NG and DC participated in critically appraising and revising the intellectual content of the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Merton RK. The Unanticipated Consequences of Purposive Social Action. *Am Sociol Rev* 1936;**1**:894-904.
2. Wachter RM. Expected and unanticipated consequences of the quality and information technology revolutions. *JAMA* 2006;**295**:2780-3. DOI:10.1001/jama.295.23.2780
3. Han YY, Carcillo JA, Venkataraman ST, Clark RSB, *et al.* Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 2005;**116**:1506-12. DOI:10.1542/peds.2005-1287
4. Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. *N Engl J Med* 2007;**357**:608-13. DOI:10.1056/NEJMs070738
5. Bardach N, Cabana M. The unintended consequences of quality improvement. *Curr Opin Pediatr* 2009;**21**:777-82. DOI: 10.1097%2FMOP.0b013e3283329937
6. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ Qual Saf* 2012;**21**:876-84. DOI: 10.1136/bmjqs-2011-000760
7. Institute of Healthcare Improvement: *Science of Improvement: Establishing Measures*. <http://www.ihl.org/resources/pages/howtoimprove/scienceofimprovementestablishingmeasures.aspx>. Accessed 21 Mar 2016.
8. Langley GJ, Nolan KM, Nolan TW, *et al.* The Improvement Guide: A Practical Approach to Enhancing Organizational Performance. San Francisco,CA:Jossey-Bass;1996.
9. Clarke J, Davidge M, James L. How-to Guide for Measurement for Improvement, *Patient Safety First* 2009. <https://eoeleadership.hee.nhs.uk/sites/default/files/Patient%20Safety%20First%20How%20To%20Guide%20measurement%20for%20improvement.pdf>. Accessed 03 Feb 2016.
10. NHS Quality Improvement Scotland: *Introduction to Measurement for Improvement*. <http://www.qihub.scot.nhs.uk/knowledge-centre/qualityimprovementtopics/measurement-for-improvement.aspx>. Accessed 21 Jan 2016.
11. Asti L, Magers JS, Keels E, *et al.* A quality improvement project to reduce length of stay for neonatal abstinence syndrome, *Pediatrics* 2015;**135**:1494-e500. DOI: 10.1542/peds.2014-1269

441 12. Mittal V, Darnell C, Walsh B, *et al.* Inpatient bronchiolitis guideline implementation and resource
442 utilization. *Pediatrics* 2014;**133**:e730-e7. DOI: 10.1542/peds.2013-2881

443 13. Turkington P, Power M, Hunt C, *et al.* There is another way: empowering frontline staff caring for
444 acutely unwell adults. *Int J Qual Health Care* 2013;**26**:71-8. DOI: 10.1093/intqhc/mzt084.

445 14. Ogrinc G, Davies L, Goodman D, *et al.* SQUIRE 2.0 (Standards for Quality Improvement Reporting
446 Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*
447 2016;**25**:986-92. DOI: 10.1136/bmjqs-2015-004411

448 15. Reed JE, Card AJ. The problem with Plan-Do-Study-Act cycles. *BMJ Qual Saf* 2016;**25**:147-52. DOI:
449 10.1136/bmjqs-2015-005076

450 16. Curnock E, Ferguson J, McKay J, *et al.* Healthcare Improvement and Rapid PDSA Cycles of
451 Change: A Realist Synthesis of the Literature. *NHS Education for Scotland* 2012.
452 http://www.nes.scot.nhs.uk/media/1389875/pdsa_realist_synthesis.pdf. Accessed 03 June 2017.

453 17. Jones EL, Lees N, Martin G, *et al.* How Well Is Quality Improvement Described in the
454 Perioperative Care Literature? A Systematic Review. *Jt Comm J Qual Patient Saf* 2016;**42**:196-206.
455 DOI: 10.1016/S1553-7250(16)42025-8

456 18. Nicolay CR, Purkayastha S, Greenhalgh A, *et al.* Systematic review of the application of quality
457 improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg*
458 2012;**99**:324-35. DOI: 10.1002/bjs.7803

459 19. Manojlovich M, Lee S, Lauseng D. A systematic review of the unintended consequences of clinical
460 interventions to reduce adverse outcomes. *J Patient Saf* 2016;**12**:173-9. DOI:
461 10.1097/PTS.0000000000000093

462 20. McDonald KM, Schultz EM, Chang C. Evaluating the State of Quality-Improvement Science
463 through Evidence Synthesis: Insights from the Closing the Quality Gap Series. *Perm J* 2013;**17**:52-61.
464 DOI: 10.7812/TPP/13-010

465 21. Pronovost P, Wachter R. Proposed standards for quality improvement research and publication:
466 one step forward and two steps back. *Qual Saf Health Care* 2006;**15**:152-3. DOI:
467 10.1136/qshc.2006.018432

468 22. Brainard J, Hunter PR. Do complexity-informed health interventions work? A scoping review.
469 *Implement Sci* 2016;**11**:127. DOI: 10.1186/s13012-016-0492-5

23. Jones, E. SIQINS: Strengthening Reporting of Quality Improvement Interventions and Methods in Surgery. <https://lra.le.ac.uk/bitstream/2381/39747/1/2017-JONES-EL-PHD.pdf>. Accessed 29 May 2017.

24. The Scottish Patient Safety Programme (SPSP). <http://www.scottishpatientsafetyprogramme.scot.nhs.uk/>. Accessed 11 May 2017.

25. Ritchie J, Lewis J, McNaughton Nicholls C, *et al*. Qualitative Research Practice. A guide for social science students and researchers. 2nd ed. London: Sage Publications; 2003.

26. Ash JS, Sittig DF, Dykstra RH, Guappone K, *et al*. Categorizing the unintended sociotechnical consequences of computerized provider order entry. *Int J Med Inform* 2007;**76**:21-7. DOI: 10.1016/j.ijmedinf.2006.05.017

27. Ash JS, Sittig DF, Poon EG, *et al*. The extent and importance of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc* 2007;**14**:415-23. DOI: 10.1197/jamia.M2373

28. Bloomrosen M, Starren J, Lorenzi NM, *et al*. Anticipating and addressing the unintended consequences of health IT and policy: a report from the AMIA 2009 Health Policy Meeting. *J Am Med Inform Assoc* 2011;**18**:82-90. DOI: 10.1136/jamia.2010.007567

29. Rogers, EM. Diffusion of Innovations. 4th ed. New York: Free Press; 1995.

30. Campbell EM, Sittig DF, Ash JS, *et al*. Types of Unintended Consequences Related to Computerized Provider Order Entry. *J Am Med Inform Assoc* 2006;**13**:547-56. DOI: 10.1197/jamia.M2042

31. Scottish Government: *Quality Improvement and Data Measurement-What Non-Executive Directors need to know*. 2016. <http://www.gov.scot/Resource/0049/00492311.pdf>. Accessed 01 May 2016.

32. U.S. Department of Health and Human Services Health Resources and Services Administration: *Performance Management and Measurement* 2011. <http://www.hrsa.gov/quality/toolbox/508pdfs/performanceandmeasurement.pdf> Accessed 09 Mar 2016.

33. Health Quality Ontario: *Measurement for Quality Improvement* 2013. <http://www.hqontario.ca/Portals/0/documents/qi/qi-measurement-primer-en.pdf>. Accessed 13 Apr 2016.

500 34. Shojania KG, Ranji SR, McDonald KM, *et al.* Effects of quality improvement strategies for type 2
501 diabetes on glycemic control: A meta-regression analysis. *JAMA* 2006;**296**:427-40. doi:
502 10.1001/jama.296.4.427

503 35. Kavanagh PL, Sprinz PG, Wolfgang TL, *et al.* Improving the Management of Vaso-Occlusive
504 Episodes in the Pediatric Emergency Department. *Pediatrics* 2015;**136**:e1016-25. DOI:
505 10.1542/peds.2014-3470

506 36. Dewan M, Galvez J, Polsky T, *et al.* Reducing Unnecessary Postoperative Complete Blood Count
507 Testing in the Pediatric Intensive Care Unit. *Perm J* 2017;**21**:16-051. DOI: 10.7812/tpp/16-051

508 37. Duvoisin G, Fischer C, Maucourt-Boulch D, *et al.* Reduction in the use of diagnostic tests in infants
509 with risk factors for early-onset neonatal sepsis does not delay antibiotic treatment. *Swiss Med Wkly*
510 2014;**144**:w13981. DOI: 10.4414/smw.2014.13981

511 38. Bell S, Davey P, Nathwani D, *et al.* Risk of AKI with Gentamicin as Surgical Prophylaxis. *J Am Soc*
512 *Nephrol* 2014;**25**:2625-32. DOI: 10.1681%2FASN.2014010035

513 39. Strom BL, Schinnar R, Aberra F, *et al.* Unintended effects of a computerized physician order entry
514 nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. *Arch Intern Med*
515 2010;**170**:1578-83. DOI: 10.1001/archinternmed.2010.324

516 40. Taylor MJ, McNicholas C, Nicolay C, *et al.* Systematic review of the application of the plan-do-
517 study-act method to improve quality in healthcare. *BMJ Qual Saf* 2014;**23**:290-8. DOI:
518 10.1136/bmjqs-2013-001862

519 41. Matosin N, Frank E, Engel M, *et al.* Negativity towards negative results: a discussion of the
520 disconnect between scientific worth and scientific culture. *Dis Model Mech* 2014;**7**:171-3. DOI:
521 10.1242/dmm.015123

522 42. Raleigh VS, Foot C. *Getting the measure of quality. Opportunities and challenges* 2010.
523 [https://www.kingsfund.org.uk/sites/files/kf/Getting-the-measure-of-quality-Veena-Raleigh-](https://www.kingsfund.org.uk/sites/files/kf/Getting-the-measure-of-quality-Veena-Raleigh-Catherine-Foot-The-Kings-Fund-January-2010.pdf)
524 [Catherine-Foot-The-Kings-Fund-January-2010.pdf](https://www.kingsfund.org.uk/sites/files/kf/Getting-the-measure-of-quality-Veena-Raleigh-Catherine-Foot-The-Kings-Fund-January-2010.pdf). Accessed 19 Jan 2016.

525 43. Dixon-Woods M, Amalberti R, Goodman S, *et al.* Problems and promises of innovation: why
526 healthcare needs to rethink its love/hate relationship with the new. *BMJ Qual Saf* 2011;**20**: i47-i51.
527 DOI: 10.1136/bmjqs.2010.046227

528 44. Dixon-Woods M, Martin G. Does quality improvement improve quality? *Future Hospital Journal*
529 2016;**3**:191-4. DOI: 10.7861/futurehosp.3-3-191

45. Vincent C, Burnett S, Carthey J. The measurement and monitoring of safety. Drawing together academic evidence and practical experience to produce a framework for safety measurement and monitoring. *Health Foundation Spotlight report* 2013.
http://www.health.org.uk/sites/health/files/TheMeasurementAndMonitoringOfSafety_fullversion.pdf. Accessed 29 July 2016.
46. Vincent C, Burnett S, Carthey J. Safety measurement and monitoring in healthcare: a framework to guide clinical teams and healthcare organisations in maintaining safety. *BMJ Qual Saf* 2014;**23**:670-7. DOI: 10.1136/bmjqs-2013-002757
47. McQuillan RF, Silver SA, Harel Z, *et al*. How to Measure and Interpret Quality Improvement Data. *Clin J Am Soc Nephrol* 2016;**11**:908-14. DOI: 10.2215%2FCJN.11511015
48. Lee SL. The Extended Surgical Time-Out: Does It Improve Quality and Prevent Wrong-Site Surgery? *Perm J* 2010;**14**:19-23.
49. Powell AE, Rushmer RK, Davies HTO. A systematic narrative review of quality improvement models in health care. *Healthcare Improvement Scotland* 2009.
http://www.healthcareimprovementscotland.org/previous_resources/hta_report/a_systematic_narrative_review.aspx. Accessed 20 March 2017.
50. Portela MC, Pronovost PJ, Woodcock T, *et al*. How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf* 2015;**24**:325-36. DOI: 10.1136/bmjqs-2014-003620
51. Goodman D, Ogrinc G, Davies L, *et al*. Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) Guidelines, V.2.0: examples of SQUIRE elements in the healthcare improvement literature. *BMJ Qual Saf* 2016;**25**:e7. DOI: 10.1136/bmjqs-2015-004480
52. Dayan M, Edwards N. Learning from Scotland's NHS. *Nuffield Trust Research report* 2017.
<https://www.nuffieldtrust.org.uk/research/learning-from-scotland-s-nhs>. Accessed 27 August 2017.

559

560 **Box 1: Institute of Healthcare Improvement recommended types of measures^{7 8}**

“Use a balanced set of measures for all improvement efforts: outcomes measures, process measures, and balancing measures.

1. **Outcome Measures:** How does the system impact the values of patients, their health and wellbeing? What are impacts on other stakeholders such as payers, employees, or the community?
2. **Process Measures:** Are the parts/steps in the system performing as planned? Are we on track in our efforts to improve the system?
3. **Balancing Measures** (looking at a system from different directions/dimensions): Are changes designed to improve one part of the system causing new problems in other parts of the system?”

561 * Adapted from IHI (text is verbatim quote but examples are omitted and text is renumbered)

562 **Table 1: Characteristics of participants in both phases of the study**

Participants' primary roles and responsibilities	Setting	Phase 1: Semi-structured interviews N=15*	Phase 2: Focus group interviews N=24**
Improvement advisors with relevant clinical background and healthcare improvement expertise both locally and nationally, external to the local clinical and managerial teams	Primary care, Maternity, neonatal and paediatrics, Mental health, Healthcare Associated Infections, High risk medicines	5	5
People with a university or similar academic base and perspective , relevant clinical background and healthcare improvement expertise both locally and nationally	Public health, Palliative care, Primary care, Community engagement, Health inequalities, Capacity and capability building	3	9
Providers of healthcare services including clinicians in leadership positions in quality and safety who retain a significant role within their routine clinical practice, being involved in delivering healthcare improvement both locally and nationally	Primary care, Mental health, Medicine for the elderly, Public health, Maternity, neonatal and paediatrics	2	5
Providers of social care services in leadership positions in quality and safety who are involved in facilitating improvements both locally and nationally across the integrated health and social care services	Community health and social care partnerships	1	1
Policy-makers and commissioners involved in monitoring performance and setting the general direction of quality improvement	Health care, social care, Education and early years	3	2
Patient representatives advising health boards on the most efficient ways of accounting for the views and experiences of the people who use the local services	Cardiac care and rehabilitation, Dementia care, Maternity care	1	2

*All interviews were conducted face to face except two academics and one policy maker which were interviewed by telephone;

**Two participants in the interviews (one improvement advisor and one academic) also attended the focus groups;

567 **Table 2: Published examples of trade-offs, pleasant and unpleasant surprises in the improvement literature**

568

Study	Improvement goals	Other consequences	Examples of balancing measures prior, during and post implementation
Kavanagh 2015 ³⁵	To improve the timeliness of management of vaso-occlusive pain events in children with sickle cell disease in paediatric emergency departments	Expected undesirable consequences (trade-offs)	<i>Mean time from triage to the second intravenous opioid dose</i> was introduced as a balancing measure because of concern that the use of intranasal fentanyl as the first-line intervention might delay subsequent intravenous dosing. Other <i>trade-offs</i> measured included <i>readmission rates</i> within 24 hours of discharge, episodes of <i>respiratory depression</i> and inpatient <i>length of stay</i> .
Dewan 2017 ³⁶	To decrease unnecessary routine complete blood count testing in a low risk cohort of postoperative patients in the paediatric intensive care units	Expected undesirable consequences (trade-offs)	Balancing measures were implemented for <i>haemoglobin level below 8 g/dL</i> in patients for whom complete blood counts were actually sent and <i>blood transfusions up to 7 days postoperatively</i> for any patients in the cohort.
Duvoisin 2014 ³⁷	To reduce the number of unnecessary diagnostic tests such as complete blood count and C-reactive protein in infants with risk factors for early-onset neonatal sepsis	Unexpected desirable consequences (pleasant surprises)	There was pre-intervention concern that reduction in the use of diagnostic tests would <i>delay the initiation of antibiotic treatment</i> , but unexpectedly the intervention resulted in <i>earlier treatment of infection</i> on average.
Bell 2014 ³⁸	To reduce the pre-operative use of antimicrobials associated with Clostridium difficile infection	Unexpected undesirable consequences (unpleasant surprises)	The new surgical prophylaxis regimen of four doses of flucloxacillin 1g plus single dose gentamicin 4mg/kg unexpectedly led to <i>increased rates of post-operative acute kidney injury</i> in orthopaedic patients, large enough to lead to the termination of the intervention through a change in the national antibiotic policy recommendation for orthopaedic surgical prophylaxis.
Strom 2010 ³⁹	To evaluate the effectiveness of a customized nearly hard stop alert in reducing concomitant orders for warfarin and trimethoprim- sulfamethoxazole compared with the standard practice of a pharmacist intervention program.	Unexpected undesirable consequences (unpleasant surprises)	Unexpected <i>delays in indicated anticoagulant and/or antimicrobial treatment initiation</i> were deemed sufficiently serious to warrant discontinuation of the improvement intervention.

569 **Figures/illustrations**

570 **Figure 1** – Draft framework of types of consequences of quality improvement projects (derived from
571 phase 1 data and the literature, sent to phase 2 participants before the focus groups)

572 **Figure 2** – Refined framework of types of consequences of quality improvement projects (derived
573 from phase 1 data and the literature, and refined after phase 2 focus groups)